11: Bootstrap Confidence Intervals with StatKey

Stat 120 | Fall 2025

Prof Amanda Luby

> These problems are drawn from the textbook with the datasets built into StatKey. I want the focus today to be on the "big ideas" of bootstrap distributions and confidence intervals, without the added complexity of coding. Next time, we'll see how to load data and create bootstrap confidence intervals in R.

Part of what you need to do is choose the right menu within StatKey. For example, if you are trying to make a confidence interval for a proportion, you will need to be in the "proportion" menu under "Bootstrap confidence intervals".

Statkey link: https://www.lock5stat.com/StatKey/index.html

1. Atlanta Commute times

What is the average commuting time for people who live and work in the Atlanta Metropolitan Area? It's not feasible to contact all Atlanta residents, but the US Census Bureau regularly collects data from carefully selected samples of residents in many different areas. This data is from the American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data represent a random sample of people who live in the Atlanta metropolitan area. They include only cases where the respondent worked somewhere other than home. Variables include the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

(a) We're going to look at the Commute Time. In StatKey, choose "Confidence Interval for Mean, Median, StDev" and then "Atlanta Commute Time". You should see the data distribution in the "Original Sample" graph. Describe the shape, center, and spread of the data distribution.
(b) What is the relevant parameter and the best point estimate?
(c) Generate 1 bootstrap sample. You should see the sample under "Bootstrap sample". How does n, mean, median, and st dev compare to the original sample?

(d) Generate 1000 bootstrap samples and describe the shape, center, and spread of the bootstrap distribution. How does it compare to the data distribution? You should be able to explain why any discrepancies exist.
(e) Construct a 95% confidence interval for the average commute time and interpret it in context

2. Compassionate Rats

In a recent study, rats showed compassion that surprised scientists. Twenty-three of the 30 rats in the study freed another trapped rat in their cage, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion. Rats did not open the cage when it was empty or when there was a stuffed animal inside, only when a fellow rat was trapped. We wish to use the sample to estimate the proportion of rats that would show empathy in this way. The data are available in the dataset CompassionateRats on StatKey.

(a) Give the relevant parameter and its best point estimate
(b) Describe how you could use slips of paper to create one bootstrap statistic. How many slips of paper do you need?
(c) Use StatKey to create a bootstrap distribution with 500 bootstrap samples. Describe the shape, center, and spread using appropriate quantities.
(d) Use StatKey to create a bootstrap distribution with 20,000 bootstrap samples. How does it compare to your distribution from (c)?
(e) Use the standard error to find and interpret a 95% confidence interval for the proportion of rats likely to show empathy.

3. Do Teen Problems Differ Based on Income Level?

A recent study found that anxiety/depression topped the list of problems teens see among their peers. The table below shows whether they consider anxiety/depression a major problem based on household income level. We are interested in estimating the difference in proportion who think it is a major problem between the two groups.

Income	Major Problem	Not Major Problem	Total
<$75,000	386	150	536
>$75,000	258	126	384
Total	644	276	920

a. What proportion of teens in households with an income below 75,000 think it is a major problem? What about the proportion above 75,000?

b. Give notation and the value of the relevant sample statistic. Give notation for the population parameter.

c. Use StatKey to find the standard error for this statistic using a bootstrap distribution. (This data is called "Teen Anxiety Depression Problem (by Income)")

d. Use the standard error to find a 99% confidence interval for the difference in proportions.

4. Atlanta Commute Times (again)

A key reason for longer commute times is having to travel a further distance. We want to see how strong the (linear) relationship is between commute distance and commute time among Atlanta residents.

a. What population parameter is of interest here?

b. Find the correct StatKey menu and choose the dataset "Commute Atlanta (Time as a function of Distance)". Describe the original sample and find the appropriate sample statistic.

c. Generate a bootstrap distribution and describe the shape, center, and spread.

d. Construct a 95% confidence interval and interpret it in context.